Identity in the London Indian diaspora: towards the quantification of qualitative data

Hundt, Marianne ; Staicov, Adina

Abstract: For second-generation members of a diaspora community, ethnic and cultural affiliation are less straightforward than for the first generation. We compare information on identity construction in London's Indian Diaspora with the participants' linguistic integration into the host community. Our study is novel and exploratory in that it combines quantitative, variationist methodology with a qualitative approach. We employ two standard sociolinguistic instruments to model subjects' ethnic identity: a questionnaire and sociolinguistic interviews with a focus on discursive identity construction. In a second step we investigate possible connections between morphosyntactic variation and ethnic identity in language use data from three different communicative contexts. The results show that, while interview data on ethnic identity are amenable to quantification, clear correlations between the resulting identity scores and vernacular morphosyntactic features are difficult to find. In particular, patterns of style-shifting between the different communicative contexts are not as expected.

DOI: https://doi.org/10.1111/weng.12311

Expressing and negotiating identity in the London Indian Diaspora: Towards the quantification of qualitative data

AUTHOR* and AUTHOR**

ABSTRACT: For second-generation members of a diaspora community, ethnic and cultural affiliation are less straight-forward than for the first generation. We compare information on identity construction in London's Indian Diaspora with the participants' linguistic integration into the host community. Our study is novel and exploratory in that it combines quantitative, variationist methodology with a qualitative approach. We employ two standard sociolinguistic instruments to model subjects' ethnic identity: a questionnaire and sociolinguistic interviews with a focus on discursive identity construction. In a second step we investigate possible connections between morphosyntactic variation and ethnic identity in language use data from three different communicative contexts. The results show that, while interview data on ethnic identity are amenable to quantification, clear correlations between the resulting identity scores and vernacular morphosyntactic features are difficult to find. In particular, patterns of style-shifting between the different communicative contexts are not as expected.

Keywords: (discursive) identity construction, linguistic variation, combining sociolinguistic methodologies

INTRODUCTION

Sociolinguistic research hardly ever combines quantitative, variationist methodology with evidence from discursive identity construction, as Mendoza-Denton remarked over a decade ago (2002: 490): '… it is relatively rare for discourse studies of identity

* address, e-mail
** address, e-mail

to have an explicitly variationist focus;' she continues that this constitutes '… a lacuna, with great potential for future research'. Similarly, Hoffman & Walker (2010: 37) point out that 'large-scale quantitative studies continue to categorize participants ethnically without much explicit discussion of how these categories are decided on or the social theory underlying these categories.' It is the aim of this paper to take a step in the direction of filling this research gap. In order to do so, we employ two standard sociolinguistic instruments, namely (a) a questionnaire eliciting information on the subjects' background (including language proficiency and occupation) and their degree of integration in both ethnic and host/mainstream[1] communities and (b) sociolinguistic interviews with a focus on discursive identity construction (e.g. Wodak et al. 1999) and meta-linguistic reflection on linguistic practices. These identity profiles are then compared to patterns of morphosyntactic variation in various speech situations. Language use data come from map tasks, the sociolinguistic interviews and discussions with another member of the community, i.e. three contexts between which subjects are likely to style-shift. The ultimate aim of our research is to correlate identity construction in a Diaspora context with the participants' linguistic integration into the host community, i.e. the degree to which ethnic variants continue to be used in various contexts.[2]

In part two of the paper, we briefly give background information on the sample used in this study. In part three, we describe the steps taken in order to quantify data from our interviews in the form of identity scores, which can then be compared with quantitative data from questionnaires. The two approaches to modeling ethnic identity are compared in section 4.1. The advantages and limitations of our approach are discussed in section 4.2. In part five, we give background on the morphosyntactic variable investigated. The results on morphosyntactic variation are presented in part

six and then mapped onto the results of the identity scores in section 6.3. In the concluding part, we critically assess the advantages of the approach taken to measuring ethnic identity and the limitations on correlation with linguistic variation. As it turns out, these limitations are – at least in part – due to the choice of communicative situations in which participants were recorded.

DATA

Our data come from a small-scale study of the Indian Diaspora in London. Participants with either Gujarati or Hindi as their first language were recruited through a contact person at University College London (UCL). They completed a map task, filled in a questionnaire, were interviewed and discussed topics amongst themselves. Typically, for the purposes of studying style-shifting in contexts of different formality, participants are asked to read out lists of words or texts to tap into a more careful register of speech. We chose the map task as a relatively formal communicative context instead, because our research design was developed with a view to also modelling ethnic identity construction and sociophonetics, more precisely the variation of speech rhythm. Since this is notoriously difficult to quantify as a variable, subjects read out parts of the map task after it had been transcribed, thus allowing for read speech normalization in the sociophonetic part of this project (for details, see Dellwo and Zipp 2011). For the study of freely occurring morphosyntactic variation, standard contexts that are employed in the study of style-shifting are not suitable, either.

The aim of the research design was to record participants in situations with different interlocutors (from their own ethnic background and with an outsider) and different degrees of 'control' (one relatively free, another very goal-oriented). The map task was completed twice by every participant: once with a member of their own

community and another time with a field worker, i.e. an outsider. Thus, in the map task and the discussion, they interact with an participant from their own community who is known to them, albeit in a relatively formal context and a more informal one. The interview and the second map task provide contexts in which they are likely to adapt to an outsider with a different ethnic background.

Our interview data are only relatively 'free' in that a standard set of questions were asked. Holliday (2010: 101) claims that this leads to an inherent weakness because such qualitative research only elicits answers to questions that the researchers thought of and is thus not fundamentally different from quantitative research. However, a closer look at the answers to the interview questions (see section 4.2) shows that open questions allow the participants to provide conflicting evidence in the same answer. Moreover, subjects may provide unexpected information (e.g. as a response to attitudes towards Indian-accented English, participants may comment on other English accents, see section 4.2.2 for an example). While these kinds of answers prove problematic for our attempts at quantification they make the research more qualitative as we can tap into our participants' own conceptualization of, and attitudes towards,ethnicity and identity, thus taking a more emic approach in our analysis tat goes beyond simply using a standard set of questions.

A total of eight participants were recruited: four each with a Gujarati and Hindi background. Our objective was to sample four first- and four second-generation participants. As it turned out, some of the first-generation participants had arrived in the UK relatively recently and it was unclear whether they intended to stay in the UK or whether they were planning to return to India after their university education. We therefore decided to include only two first-generation participants in our study (i.e.

leaving a total of six participants). Table 1 gives detailed information on the composition of our sample, including the relationship between the two participants.

For the variationist part of the study, H04 and G02 were not included because first-generation participants typically have access to a more restricted repertoire and are thus likely to style-shift less with different interlocutors than members of the second-generation.

[Table 1 here]

Of the second-generation participants included in the variationist part of the study, H01, G01 and G03 all had interlocutors from the second generation, whereas G04 interacted with her mother, a first-generation informant. It might be the case that a second-generation participant accommodates more to a first-generation participant in the more informal context (i.e. the discussion).

This paper presents a case study that zooms in on the linguistic behavior of young, second-generation British Asians, and the small number of participants does not allow us to generalise our findings to the speech community in London. Nevertheless, our results reveal tendencies of how identity, style and interlocutor interact with linguistic practices, which can serve as stepping stones for further exploration of how British Asians use their linguistic repertoire for indexing ethnic identity. In particular, they show that the methodology we adopt provides a novel way of modelling identity construction.

METHODOLOGY: MODELLING IDENTITY CONSTRUCTION

As was typical of first-wave, quantificational sociolinguistic research, ethnic identity (like other social categories) was taken to be a fixed and stable variable that was applicable across a broad range of speech communities (see Schleef, forthcoming). Consequently, early sociolinguistic studies on language and ethnicity

did not reflect the multifaceted nature of an individual's identity. The past few decades have seen a shift from this essentialist view towards a more constructivist approach to identity and ethnicity that recognizes identification and affiliation to an ethnic group to be fluid and subject to situational change (Bell 2002; Joseph 2004; Bucholtz and Hall 2004, 2005). We follow this constructivist approach and employ a methodology that aims to gain an understanding of ethnic identity as it emerges from local conceptualization, i.e. as a socially meaningful dimension for the individuals in the community. Information on identity construction in our study comes from two sources: answers to questions in the interview part of the data collection as well as those that stem from a questionnaire, i.e. a fixed set of question items with Likert scales. The interviews provided the opportunity for more discursive identity construction that still needed to be quantified. There is no established methodology for the quantification of such data. In the next section we therefore outline the steps taken to achieve this.

Quantifying interview responses

In a first step, we listened to the interviews and took notes on the themes that were covered. As pointed out in the previous section, we asked a diverse set of questions in the interview with a view to ensuring comparability while at the same time providing speakers with many opportunities to construct their ethnic identity. The topics in the interview thus serve a double function: ensuring comparability across interviews and providing potential stepping stones to further topics that were not initially included in the interview design. The questions concerned participants' place of residence, friends, bilingualism, maintenance of transnational ties and attitudes towards their country of origin, attitude towards Indian-accented English, experiences of being 'othered', etc. As discussed in section 2, the inclusion of these different topics allows

for a more nuanced assessment of ethnic identity. Taking issues like transnational ties into account, whether practiced or symbolic (Espiritu & Tran 2002; Duff 2015), further allows for the investigation of the role the ancestral home plays in our speaker identity (re-)construction.

Spolsky (2016: 102) points out that '[l]anguage, whether its absence, or "incorrectness", or accent, or style, is also a major identifying feature" in diaspora communities. The interviews therefore included various aspects of language use, including bilingualism. When discussing bilingualism, speakers not only describe their own competence in the heritage language, but also have the opportunity to express what role the heritage language plays for identity construction or whether they are competent speakers or not. Maintenance of the heritage language, similarly to the maintenance of transnational ties, can take a symbolic function that is likely to strongly affect ethnic identity construction (see AUTHOR 2016).

There was a lot of overlap between the interviews with first- and second-generation participants as the data coding sheets in appendix 1 show. In a second step, both analysts listened to all interviews again and recorded the information provided by the participants on each of the various themes. The answers were assigned scores ranging from 1 (ethnic) to 5 (host). For each theme, the mean was calculated from the individual responses. An participants' overall identity score was then calculated as a mean of responses across themes in the data coding sheet. We then compared our results for inter-annotator agreement. As it turned out, the identity scores agreed closely in some instances (participants H01, H04 and G04) but diverged for others (G02 and G03). The reasons for divergence will be discussed in section 4.2 – the discussion allows us to refine our methodology for future research. In a final step, we compared every coding sheet in detail and thus arrived at an agreed rating.

Quantifying questionnaire responses

In addition to the qualitative data from the interviews, a questionnaire was designed to elicit information on the subjects' personal and linguistic background, including information on their parents' first languages. In part two of the questionnaire, speakers were asked to assess their proficiency in both the heritage and the mainstream language. The subjects rated their proficiency in speaking, understanding, reading and writing the respective language on a seven-point scale ranging from *not at all* (1) to *very well* (7).

Part three of the questionnaire consisted of questions related to subjects' ethnic identity. For quantification of this part, we adapted the revised Multigroup Ethnic Identity Measure (MEIM-R) designed by Phinney (1992, revised in Phinney and Ong: 2007). The MEIM-R consists of six items (see Table 2) that can be rated on a scale from *strongly disagree* (1) to *strongly agree* (5). The score is calculated as the mean of all items.[3]

[Table 2 here]

Finally, the last questionnaire item aims at measuring the speakers' integration into ethnic and non-ethnic networks by asking about relationships with Asian and non-Asian peers. The answers can range from *mainly non-Asian* (1) to *mainly Asian* (3), with (2) accounting for a mix of both.

IDENTITY SCORES

Comparing interview and questionnaire ratings

Table 3 gives an overview of the results obtained from the initial identity scores by the two analysts, followed by the agreed rating and the mean calculated from the questionnaire data.

[Table 3 here]

As far as different kinds of data (questionnaire vs. interview) are concerned, the identity scores we obtained match quite well. The one exception is H04, a first-generation participant whose first language is Hindi. We will take up this case in the discussion (section 4.2).[4]

With respect to ethnic vs. host identities, none of our participants appear close to the extreme ends of the scale. As far as first- and second-generation participants are concerned, the quantitative results from our study show that they do not fall into two distinct groups but seem to form a gradient: on the basis of the quantified interview and the questionnaire data, H01 and G04 – both second-generation Indians – are quite close to G02, a first-generation informant, whereas G03 (second-generation) appears to have a slightly more 'mixed' identity.

Discussion

Problems in quantifying qualitative data from interviews

One of the reasons for divergent ratings obtained in our initial coding was that not all utterances were allocated to one of the rating themes, i.e. annotators recorded answers with different degrees of detail. The question that follows from this is whether it is legitimate to summarize answers in the ratings. This is probably the case if the trend of a set of answers goes into the same direction (as in the two answers given in (1) but proves problematic if an answer combines two aspects (see (2) and (3); all examples are from the interview with G03, a second-generation female participant whose first language is Gujarati.

1) a. would not tell parents about boyfriend only if ready to get married
   b. not acceptable to have a boyfriend

2) a. I would like my children to be able to speak Gujarati
   b. BUT: I do not think that my children will be able to speak it

3) a. participant speaks Gujarati with parents (she rarely speaks to her mother in English)

b. BUT: Whole family speaks English (so parents speak it more now)

To account for such conflicting answers, all utterances and ratings across speakers were collected into one data sheet to ensure the highest possible consistency of ratings. Ideally, full transcription of the interview data would be a helpful – if somewhat labour-intensive – initial step, especially for a larger study. The relevant parts of the interview could then be extracted with the help of keywords from concordances if the latter provide enough context to ensure rich coding of data.

At times it proved difficult to assess an utterance in terms of its identity value. The annotation in (4) by the analysts can be used to illustrate this:

4) a. East London (Newham?), Pakistani, Bangladeshi – ethnic minorities the majority; traditional views – is why Asian communities stick together [annotator 1]

b. Asian communities tend to stick together; very gossipy; community cling to culture from back home [annotator 2]

Both analysts picked up on the comments 'traditional views' and 'stick together', but out of context these are not obvious with respect to the value they should take in assessing the subject's identity construction with respect to the migrant community. A closer look at the context (annotator 2) shows that, in the interview, 'traditional views' and 'stick together' are given relatively negative values because they co-occur with value judgements such as 'gossipy' and 'cling' to culture. This negative evaluation, in turn, means that the comment cannot be interpreted in terms of positive identification with the ethnic background.

The added value of qualitative data

Both the quantified data from the sociolinguistic interviews and the questionnaire data show that the differences between the first and the second generation are not very

large. Even though this seems to be a surprising result at first, it fits in with what Williams (2005: 2352) discovered in her study of identity construction among Chinese Americans in the San Francisco Bay area: 'Examining the questionnaire responses on ethnic identity and community practices likewise indicates that the two generations are not, at least quantitatively, very different.' The interview data thus add an important dimension to the analysis. As Williams (2005: 2355) points out, they 'can capture alterations, beliefs, and fluidity, but not the general patterns.'

Participant G03, for instance, has a relatively mixed identity compared with other second-generation speakers. An utterance (not included in our quantification) that underscores this rating is given in (5) where G03 stereotypes and 'others' *freshies*, i.e. recently arrived Indian immigrants to the UK:

5) I wouldn't associate myself with that as … like that's not someone I would have as a friend … not somebody I would date I would consider myself slightly different … even though I guess we did have the same background but we've had very different upbringings and very different interests but I wouldn't say I'm better than them just different

In other words, if quantified, interview data may well fit into the general pattern, but the qualitative data enrich the analysis and allow researchers to keep track of the underlying layer of 'alternations' and 'fluidity'. Participant G01 is a particularly good example:

- She at first claims that she is (originally) from a part of London that is 52% Gujarati, only to mention later in the interview that she moved out and now lives in a flat in Camden.
- She evaluates the fact that she is able to speak both Gujarati and English positively but also points out that, as a child, she resented having to go to Gujarati school.
- She admits that it 'bothers her' to hear somebody speak with an Indian accent but at the same time she says that she 'hates it' when Indians try to put on an American accent.

In other words, the more open-ended nature of questions in the interview context allow participants to provide information that, at times, may be contradictory.

The participant showing the greatest divergence between the identity score obtained from the interview and the result of the questionnaire is H04, a first-generation male. He scores the highest rate for 'ethnic' identity in the questionnaire but has a fairly 'integrated' ethnicity according to the interview data. A closer look at his answers reveals that it is important to him not to stand out: H04 describes how after his family's arrival in the UK, fellow students mocked his accent and how it was difficult for him to understand the British accent and to communicate with teachers and students. While he did not actively try to change his accent, he felt that he needed to improve his English and that, over time his accent changed by itself; he now feels that he has adapted quite well to British culture. It may well be the case that in answering the questionnaire, he felt more confident in asserting his ethnic identity and 'accommodated' towards the researcher by down-toning his ethnicity in the interview setting. Participant H04 may thus be a good example of how much questionnaire and interview data may occasionally diverge. This is not entirely surprising if we consider De Fina (2011: 30), who points out that identity construction in an interview setting "is a dialogic and reciprocal process in which 'the other' is as important as the self; the kinds of identities that people present crucially depend on who they understand the interlocutors to be." For individual participants, this might mean that they amplify their host-community orientation with an interviewer who is not from the community.

LINGUISTIC VARIATION

In order to investigate how the identity scores obtained in the first part of our study would map onto patterns of linguistic variation, we chose to focus on variable article use and other vernacular morphosyntactic features. Omission of articles is a pervasive feature of various varieties of Indian English (see for instance Agnihotri et al. 1994 or Sharma 2005). Among other vernacular features, non-standard use of prepositions

was particularly wide-spread (partly due to one of the contexts of data collection, i.e. the map task). In this section, we define the linguistic variables investigated and illustrate them with examples from our data.

Variable article use

The use of definite and indefinite articles is typically variable in varieties of Indian English in contexts where speakers of English as a Native Language (ENL varieties) require either a definite or an indefinite article. Examples of a definite and an indefinite null article from our data are given in (6) and (7), respectively:[5]

6) even if they were from the same religion Ø *same community* I still wouldn't tell them (G03_I)

7) a. ... we're not going as *Ø delegation* (G03_D)
b. when you're […] standing at the bus stop and you see *Ø Indian lady* and go up (H01_I)

We coded for such null articles manually by analyzing the transcripts of the map task, interview and discussion data. Contexts in which articles are optional in ENL varieties were excluded from the counts. This is the case, for instance, with certain plural NPs, as in *(The) students were late for class* (see Hundt, 2014).

In the analysis of null articles, some instances are problematic. In the following, we discuss some of these and point out how we treated them in our analysis. In example (8), for instance, *Taj Mahal* is a proper name, and these are typically not preceded by a definite article in English. However, in this particular instance, both a preposition and a definite article would be required in standard BrE, and we therefore included (8) among the set of null articles.

8) I think I really wanna go back and go to the places I've not been like Ø Taj Mahal and stuff (G01_I)

Crucially, examples (6)). and (8)). were uttered without audible pauses immediately before the null article.

Since we are dealing with spontaneous spoken language, instances where speakers change construction mid-sentence are also potentially difficult to classify. Examples (9) and 10) illustrate this kind of problem. In (9)), the speaker uses a possessive pronoun which would have occupied the determiner slot in a noun phrase headed by *age* but then decides to rephrase the utterance using an adjective as pre-modification; this rephrasing would have called for the use of a definite article in standard BrE, turning (9) into an instance of null article usage.

9) my cousins are my/ Ø same age as me (G01_I)

In (10), the speaker uses an indefinite article for the first noun, decides to rephrase the utterance and uses a noun phrase with a pre-modifying adjective but without the indefinite article required in standard BrE; example (10) was therefore also included among the null articles.

10) so it's like a mixture Ø hybrid language yeah (G01_I)

A particular problem in spoken language is the fact that utterances may be elliptical; turn-initially, for instance, articles may be left out in colloquial varieties of BrE. For this reason, the missing article in (11) was coded as 'unclear' (and thus excluded from the variable contexts) rather than counted as a null article.

11) festival is on Saturday and Sun Sunday (G03_D)

Other examples that were excluded from the statistics were those with an unclear head noun (see example (12) as it was impossible to decide whether this head noun required the use of an article or not.

12) so it's arranged but it's a bit of a modern <unclear>word</unclear> … (G03_I)

In addition to instances of null articles we also coded instances of hypercorrect article insertion, i.e. contexts where ENL varieties would not use a definite or

indefinite article, as in examples (13) and (14) (the latter being an interesting instance of self-correction):

13) there was a Island in the middle of the ocean and that's where *the* Shutter Island was shot (H01_D)

14) they asked uhm just our ethnic and uhm where our parents are from and uhm what language no what *a* language they speak (H01_D)

As it turned out, hypercorrect articles were really infrequent (at a total of 5 and 6 instances in the interview and discussion data, respectively). They will therefore not be included in the discussion on variable article use. The fact that second-generation participants (like G01 and G03) also use hypercorrect articles lends further proof to the fact that article use remains challenging for second-generation speakers.

Indian English vernacular features

In addition to variable article use, we also collected evidence on the presence of other morphosyntactic features that are typical of Indian English (see e.g. Sedlatschek 2009, Bhatt 2008). As with variable articles, data collection for vernacular features in our data was done manually by close reading of the transcribed map tasks, interviews and discussions. Among the features we coded for are instances of double comparatives (example (15), of demonstrative *that* used instead of the definite article (as in example (15)); see Sharma 2012) invariant tag *isn't it* (see examples (16) and (17)) and zero past tense forms (see examples (18) and (19)).

15) so yeah my dad speaks *a bit more better* English than my mum (G03_I)

16) I had five exams so I couldn't do both at the same time whereas uhm *that* time will be quite good (G03_D)

17) yeah you did gymnastics *isn't it* (G01_D)

18) oh yes we've got the same thing *isn't it* (G01_D)

19) I was like he didn't even wish me happy birthday and then she *text* my dad and said […] (G01_D)

Occasionally, it is difficult to distinguish between more generally used vernacular features and features typical of the diaspora community. In colloquial speech, the adjective *proper,* for example, can be used as an adverb pre-modifying an adjective (see (20)).[6] Some of our participants extended the use to preverbal modification (as in example (21)).

20) it was so much fun with her do you remember I was like *proper* fine with her (H01_D)

21) I mean like <?>even</?> before I *proper* used to trust him and everything (H01_D)

Initially, we thought we would include the latter among our examples of vernacular morphosyntax. On closer inspection, we found it to be used in the North of England, e.g. Lancaster and in Geordie.[7] And even if it is not widespread in London English, the attestation of this feature in other varieties of British English means that we had to exclude this example from our analysis.

It lies in the nature of the map task that subjects frequently have to use prepositions to direct their interlocutors along the path indicated on their map. Examples of some standard prepositions used in the map tasks are given in (22) and (22).

22) turn left and the fields should be *below* your <mhm> path (G01_MT)

23) you have to go *down* and *around* treasure buried here (G03_MT)

But our participants also often used atypical prepositions, e.g. *underneath of* for standard English *below, on top of* instead of the expected *above,* or temporal *before* and *after* instead of spatial *in front of* and *behind* (see examples (24) to (27)).

24) go from sort of *underneath of* the telephone box <mhm> turn left (G01_MT)

25) and just where, *on top of* the chestnut tree is the finish (G02_MT)

26) but go just *before* the phone box (G04_MT)

27) so do you turn *after* the hut or <unclear>words<unclear> past the telephone box (G02_MT)

We also included instances where no preposition was used among the instances of vernacular morphosyntax; examples from the map task data of such null prepositions include examples (28) to (30) .

28) so I'm on top Ø the hill (G04_MT)

29) shall I go Ø 'treasure buried here'[8] (G04_MT)

30) so walk , … uhm r- across … across Ø right (H01_MT)

Another vernacular feature of Indian English is the use of progressives in non-progressive contexts (see e.g. Rogers 2002). Instead of using a (non-progressive) imperative verb form to direct their partners through the map task, some of our participants regularly opt for the progressive, as illustrated in (31) and (32):

31) you *should be passing* Ø the left of it (G01_MT)

32) so *you're going you're going* around the lorry (G03_MT)

While variable article use could be measured in terms of the proportion of null articles against definite/indefinite article (i.e. as a choice context), other vernacular features in our data are given as normalized frequencies (i.e. relative to the number of words produced by the informant).

MORPHOSYNTACTIC VARIATION

In this section, we first present the results on null articles and the overall use of vernacular features in the map task, the interview and the discussion. In a third step, we investigate how the results on language use and style-shifting map onto the results from part one of the paper, i.e. the ethnic identity scores obtained from the combined questionnaire and interview data.

Null articles

Before we present the results obtained from the second-generation participants, we will take a look at article use in one first-generation participant (G04_D) whose article use will be taken as a baseline.[9]

[Table 4 here]

Table 4 shows that the baseline speaker uses significantly more null articles in the informal context (discussion) than in the more formal map task. In a chi-square test, the difference proves significant at $p = 0.002$, 2df (corrected for continuity). This indicates that null articles may be a highly context-sensitive feature.

The results of variable article use for our second-generation participants are given in Table 5, separately for in- and out-group interlocutor for the map task (MT).

[Table 5 here]

Contrary to what we expected, null articles are overall more frequent in the map task (the most formal context) than in the interview or the discussion (the subsequently less formal contexts). This may have something to do with the fact that a map task is not a communicative situation that participants typically engage in; they may therefore have focused on the task instead of the formality of the situation. Overall, null articles are used more often with an in-group than an out-group interlocutor in the map task (with the exception of G01, whose rate of null articles increases slightly from 8.7 to 10.6). Interestingly, H01 has a null article ratio similar to our baseline speaker in the map task setting with an in-group interlocutor. In other words, some participants seem to style-shift in their use of articles between in-group and out-group communication (but see section 6.3 for a counter-example).

If speakers were generally adjusting their null article use to the formality of the situation, we would expect the ratio of null articles to increase from the interview to

the discussion (i.e. with a decrease in formality of the situation). This is not the case for most of our participants: H01 and G01 actually use fewer null articles in the discussion and G03's ratio remains the same in both contexts. It is only G04 whose proportion of null articles increases dramatically from the interview to the discussion. This is interesting because G04 is the only participant interacting with a member from the first generation (her mother), whose null article use increases significantly to 45.2% in the discussion (see Table 4). In other words, the second-generation participant is style-shifting to a markedly more vernacular style in the discussion whereas in the interview, null articles are completely absent. Whether these patterns of null-article use correlate in any way with identity scores will be discussed in section 6.3.

Vernacular features

Again, we start presenting our results by giving information on the usage of vernacular features (null articles included) in our first-generation informant, who provides the baseline for comparison. This time, there is no significant difference between the map task and the discussion (Table 6). There is no skewing towards a particular type of vernacular feature in the two contexts.

[Table 6 here]

Table 7 gives an overview of the frequency with which the second-generation participants use vernacular features across the different contexts.[10]

[Table 7 here]

In the map task, G01 and G04 seem to follow the expected pattern: both shift markedly to a less ethnic style in the out-group setting. H01 and G03, on the other hand, use a higher ratio of vernacular features with the out-group interlocutor, with H01's ratio being very close to that of the first-generation speaker.

Overall, our second-generation speakers use markedly fewer vernacular features in both the interview and the discussion than the base-line first generation speaker. With no relative frequency higher than 4.1 in the interview and 5.4 in the discussion, the difference to 47.7 per cent vernacular density of the base-line speaker is very striking.

The frequencies for the interview and discussion data follow the expected pattern for H01 and G04, with both speakers using a less ethnic style in the interview setting. G03, on the other hand, is barely influenced by the different interlocutors in the interview and discussion while G01 shows a markedly different behaviour and employs a markedly more ethnic style in the interview setting. On closer inspection, the vernacular features in G01's interview situation are mostly null articles, a feature that previous studies (e.g. Sharma 2005, Hundt 2014) have shown to be very persistent due to typological differences between English and non-article languages such as Hindi or Gujarati. Table 8 therefore gives the results for vernacular features excluding null articles.

[Table 8 here]

In the map task, excluding the articles did not change the tendencies previously observed. H01 style-shifts between interview and discussion in the way described earlier; G01 does not markedly style-shift between the situations once null articles are excluded (as we would expect from what was observed above). Similarly, no significant style-shifting can be found in G03's behaviour. We saw above that G04 is the only participant whose use of null articles correlates highly with interlocutor in all contexts. Once null articles are removed from her counts, she no longer accommodates as clearly to the in-group interlocutor in the discussion.

The findings in the present study are substantiated by Nordmann's (2016) investigation of rhoticity in the same sample of speakers analysed here. Her results show that with regard to variation across generations, second-generation speakers pattern more closely to the host community than the first-generation speakers.

Mapping linguistic variation onto identity scores

In Table 9, we repeat the ethnic identity scores for the participants whose usage data we investigated, adding a combined score (mean) from the interview and questionnaire data for ease of reference.

[Table 9 here]

When we look at the patterns found for use of null articles, it is difficult to see a clear correlation with ethnic identity. Overall, G04 is leaning towards a 'host' ethnic identity with scores at 2.4 and 2, yet she is the only participant who style-shifts markedly towards higher null article ratios with in-group interlocutors. G01, on the other hand, has a genuinely hybrid identity (with scores ranging from 2.5 to 2.8), but uses more null articles with out-group than with in-group interlocutors. This may reflect the informant's ability to control for this feature under pressure (map task) rather than it being used as a feature that marks a hybrid ethnic identity. Similarly, H01 – an participant with a relatively strong 'host' ethnic identity according to the questionnaire data – does not show a clear correlation between article use and style-shifting: the lowest rate of null articles for this participant comes from the informal discussion with an in-group interlocutor. Null articles are thus unlikely to be a marker of ethnic identity.

Let us now turn to possible correlations between other vernacular features and identity scores. Table 10 provides the frequency of vernacular features across different contexts. For clarity, the same data is given in bar chart in Figure 1.

[Table 10 here]

[Figure 1 here]

Contrary to her article usage, H01 does exhibit a higher use of vernacular features in the discussion with an in-group interlocutor than in the interview with an out-group interlocutor. For the map task, however, the tendency is reversed and H01 uses more vernacular features in the out-group setting. Overall, she has the highest ratios across all speech settings, despite her identity score leaning towards a 'host' community identity. G01 and G03 both have ethnic identity scores around the middle of the scale, suggesting 'mixed' identities, yet both speakers only show little style-shifting between the interview and the discussion setting, with G01 having a slightly higher ratio with the out-group interlocutor, and G03 exhibiting almost equal ratios in the interview and discussion settings. For the map task, G01 again uses fewer features in the out-group setting, reversing the tendency she showed with regard to article use in the same context. G03, on the other hand, uses markedly more vernacular features with the out-group interlocutor, which is contrary to her behaviour as regards variable article use.

As mentioned above, G04 presents a somewhat different case as she was recorded in an interaction with a first-generation speaker as in-group interlocutor. Her identity score leans towards a more 'host' identity and, as with the null articles, she style-shifts noticeably in her use of vernacular features, this time, however more so in the map task setting. In the less controlled interview and discussion setting, her style-shifting is less pronounced and (like G01), G04 has a slightly higher frequency of vernacular features in the interview with an out-group interlocutor. This behaviour is also described in AUTHOR (in preparation) and might suggest that a vernacular

feature is used for overt indexing of identity in the out-group setting while this is perceived unnecessary in the in-group setting.

To sum up, a straight-forward correlation between ethnic identity and vernacular features cannot be observed (in the sense that participants with a 'more' ethnic identity generally tend to use more vernacular features). This patter is in accordance with AUTHOR's (2016) and AUTHOR's (in preparation) on Chinese Americans, who also found that while different degrees of style-shifting can be observed for all speakers between the different speech settings, there is no clear correlation with ethnic identity. Results on style-shifting are complicated by our choice of the map task as the most formal communicative situation.

CONCLUSION

This paper has shown that it is possible to quantify information from sociolinguistic interviews. Moreover, our results demonstrate that the identity scores obtained in this way closely match those based on responses elicited with the help of questionnaires. The advantage of combining the two methodologies is that the interview and the questionnaire data provide the researchers with two different ways of measuring ethnic identity construction. The added value of the interview data lies in the fact that the participants are in a position to give more nuanced answers that reflect the complexity of identity construction and negotiation in a multilingual urban context and the fluidity of these identities. These, in turn, prove useful for the interpretation of the patterns of linguistic variation across different contexts and with different interlocutors. Furthermore, the interview data allow us to assess the answers given in the questionnaire, e.g. with respect to language proficiency. The first-generation participant H04, for instance, rates his current proficiency in understanding English as quite high. Looking at the statements from the interview, however, we learn that

during the first couple of years in the UK, he had difficulties understanding the British accent, information that would have been lost, had we only employed the questionnaire. To be able to account for the importance of individual topics such as language proficiency, language maintenance or transnational ties, future research may include e.g. principal component analysis (see Hoffmann & Walker 2011, AUTHOR in preparation) as this enables the calculation of the strength individual themes have on identity construction.

For second-generation members of a diaspora community, ethnic and cultural affiliation are less straight-forward than for their parents' generation. Furthermore, few studies so far have investigated variation among second-generation speakers in relation to the degree of linguistic integration into the heritage or host communities (but see Sharma 2011, Sharma & Sankaram 2011). Finally, depending on the potentially different socio-cultural and socio-economic scenarios encountered at time of arrival or growing up, we suggest that future research should use data from first-generation participants and both older and younger second-generation participants (see Sharma 2011; AUTHOR in preparation).

As regards morphosyntactic variation, our data show that speakers engage in a certain amount of style-shifting, both with regard to variable article use as well as with respect to usage of vernacular features. Two caveats apply. First, it is important to remember that this study is based on rather small samples and individual uses can thus have a significant effect on the article and vernacular density measures. Second, it turned out that the map task – which was chosen to enable speech rhythm normalisation – is probably not an ideal communicative situation to study style-shifting in the use of morphosyntactic variables: subjects might find the task itself cognitively challenging and therefore are probably unable to pay as much attention to

how they speak as to what they are saying. This may explain that, while the task is quite formal, we observe relatively high proportions of vernacular features.

With respect to inter-generational patterns of variation, our results fit in with previous research. AUTHOR (2016) and AUTHOR (in preparation) investigated speech rhythm and morphosyntactic variation, respectively, in the San Francisco Chinatown diaspora community. While variation exists in second-generation Chinese Americans, both studies find an overall tendency for second-generation speakers to behave more similarly to the mainstream community, especially with regard to younger speakers.

While our study has shown that quantification of identity scores from interview data is feasible, the set of communicative contexts obviously limits the possibility to show correlation with style-shifting. Future studies should combine interview data with self-recorded, spontaneous interaction alongside interview and discussion data. These recordings could then be used to establish more detailed linguistic ethnolinguistic repertoires for a sub-set of the participants. Research by Sharma (2011) has shown that such data are necessary to supplement Labovian-type analyses and thus arrive at a more accurate interpretation of speakers' style-shifting behaviour.

ACKNOWLEDGEMENT

NOTES (see below)

[1] While for the first generation, the label 'host' is clearly appropriate, the second generation are quite likely to feel at home where they were born, and the label 'mainstream' (as opposed to heritage) culture may therefore be more fitting. In the following, however, we use both terms interchangeably.

[2] A related project is AUTHOR (in preparation), who looks at identity construction and morphosyntactic variation in the San Francisco Chinese Diaspora community.

[3] In the pilot study, we kept two items (*I participate in cultural practices of my own group, such as holidays or religious events, music, or customs* and *I am active in cultural organizations*) from the original MEIM (1992). However, we did not use the ratings for our calculation of ethnic identity scores, as we felt that the topics were better assessed in the interviews.

[4] AUTHOR (in preparation) applies the same methodology to a larger sample (in this case the Chinese American community in San Francisco Chinatown). This enables her to measure dispersion and evaluate the differences between the groups of participants statistically with a t-test.

[5] All interviews were transcribed orthographically but not normalized in any way (e.g. forms like *innit* would be included as variant forms). The letters following speaker identification specify the context from which the examples are taken, i.e. I for 'interview', D for 'discussion' and MT for 'map task'.

[6] The *Urban Dictionary* online provides the following examples of this colloquial use of *proper*:

a. I was proper tired after spending the day out.

b. 'I'm proper hungery [sic!]'

c. 'This food is proper good'.

Pre-verbal use of *proper* is attested in the North East of England, see Fehringer and Corrigan (2015: 211).

[7] See for instance the following YouTube video clip at https://www.youtube.com/watch?v=UCOC1YwNwZw (Katherine Williams, p.c.).

[8] This was a label given on the map that the participant is quoting. Note that omission of *to* is also a feature of Multicultural London English (MLE, see e.g. Cheshire et al. 2011 or Fox 2015) and that distinguishing between features of this variety and more narrowly defined 'ethnic' features is not always possible.

[9] There are no interview data for our baseline speaker because only second-generation participants were interviewed.

[10] Appendix 2 provides information on corpus size.

[11] The identity indices calculated for the interview data, 1 (*ethnic*) to 5 (*host*), were converted to allow for comparison with the MEIM-R index,1 (*strongly disagree* = *host*) to 5 (*strongly agree* = *ethnic*).

APPENDIX 1: SAMPLE CODING

Second generation G04

Theme	Informant's comments	Questionnaire	Rating
neighbourhood/place of residence	Harrow : Asian area but her street is not v Asian (a lot of British people); family (paternal) live close		1 3 1 **1,6**
friends (1) - childhood	a group of white		4

	friends (mixed); at A-level (maths, science) changed to mainly Asian friends		2 **3**
friends (2) - university	mostly Asian, but mostly non-Gujarati speaking – more shared environments to meet these friends in; most friends are second-generation more to talk about	Friends: 3	2 2 1 **1,6**
reactions to experience of being 'othered' by host community members	Positive: she felt good about being able to wear a Sari to school on special occasions some think she's from Pakistan: when people are mistaken she corrects them		no rating 1 **1**
transnational ties (1) – type (family vs. other)	family (paternal) in Gujerat fathers family		
transnational ties (2) – frequency, length, current relevance	has only been three times (4, 11, 16) but for long periods (12 weeks)		3 **3**
transnational ties (3) - experience	likes it, but tends to get bored because of lack of internet; did voluntary work last time to avoid getting bored; would not like to go more often (see above) education in diaspora is different –		3 2 4

	tendency for relatives in India to get married earlier		4 **3,25**
bilingualism (1) – attitude, including question of whether to pass it on to the next generation	Would like to pass on Gujarati to their children, even if only passive knowledge; happy to speak Gujarati they should know it's part of the culture		3 1 1 **1,6**
bilingualism (2) – proficiency, range of use	English named first, then Gujarati Gujarati at home and in Gujarati school Often speaks English back to her mother because she is more fluent in English than in Gujarati, but tries to make the effort (couldn't do map task in Gujarati) Speaks English with her father Speaks Gujarati when she's in India (aunts and uncles have no knowledge of English) Speaks English with her sister; English easier/faster used Gujarati with some friends in school now rarely with two friends who are Gujarati as well	First at home: English, Gujarati Parents as children: Gujarati Heritage language: Gujarati Proficiency: Speaking 6, Understanding 5, Reading 7, Writing 7	3 1 4 5 1 5 5 2 3 **3,2**

attitude to Indian-accented English – 'othering' of *freshies*?	Doesn't think differently of people with a strong Indian accent. (Sees similarity with family members who have a similar accent.) can tell if someone is first generation there's a difference but they speak good English		1 4 3 2 **2,5**
cultural affiliation (music, film, clothes, festivals, membership in organization)	food, sense of community; attends festivals; is a recent developments (see friends); is into Indian films (but needs subtitles because they are in Hindi); learnt (about) Indian music and other cultural background at Gujarati school sees people who come from same background regularly community events would not wear traditional dresses except for cultural events	Finding out about ethnic group: 4 Sense of belonging: 5 Ethnic group membership: 4 Understand ethnic background: 4 Talk to other members of ethnic background: 4 Strong attachment: 3 Participation in cultural practices: 5 Active in organizations: 5 MEIM: **4,25**	2 2 2 1 1 4 **2**
generational differences in cultural values?	parents relatively relaxed in comparison to other		3

	people's parents; parents understanding and thinking is different her parents prefer for their children to finish their education before they get married;		3 3 **3**
labeling (ethnic, mixed, host)	British-Asian (is fine to be in same category as people from Pakistan and Bangladesh)	Indian	3 **3**
other	Can tell the difference between first- and second-generation migrants ('instinct') – even if their English is really good. Does not know what to answer as response to question whether she could use English to express her Indian identity. (whole concept seems strange)		
	Characteristic quote that sums up informant's identity? I have something different about me 'I'm from here.' 'What's my Indian identity?' **Interview: 2.428240741 (2,4)** **Questionnaire MEIM-R: 4**		

APPENDIX 2: CORPUS SIZE

Table 5a-1. Corpus size (word count per speaker per communicative context)

Speaker	MT_IN	MT_OUT	Discussion	Interview	Total
G01	606	461	1522	1970	4559
G03	1620	611	2004	5455	9690
G04	1246	461	2179	2946	6832
H01	1622	648	1845	2747	6862
Total	5094	2181	7550	13’118	27’943
G04_baseline	-	-	1489	-	
Total	5094	2181	9030	13’118	29’432

REFERENCES

AUTHOR. In preparation. *Narrative Identity Construction of First and Second Generation Chinese Americans Living in San Francisco Chinatown*.

Agnihotri, R. K., A. L. Khanna &Aditi Mukherjee. 1994. The use of articles in Indian English: Errors and pedagogical implications. In R. K. Agnihotri and A. L. Khanna, (eds.), *Second Language Acquisition: Sociocultural and Linguistic Aspects of English in India,* 178–198. New Delhi, India: Sage Publications.

Bell, Allan. 2002. Back in style: Reworking audience design. In Penelope Eckert & John R. Rickford (eds), *Style and sociolinguistic variation,* 139–169. Cambridge: Cambridge University Press.

Bhatt, Rakesh M. 2008. Indian English: Syntax. In Rajend Mesthrie, (ed.), *Varieties of English, Volume 4: Africa, South and Southeast Asia* (Varieties of English around the World), 546–562. Berlin: de Gruyter.

Bucholtz, Mary & Kira Hall. 2004. Language and identity. In Alessandro Duranti (ed), *A companion to linguistic anthropology,* 369–394. Malden: Blackwell.

Bucholtz, Mary and Kira Hall. 2005. Identity in interaction: A sociocultural linguistic

approach. *Discourse Studies* 7.4-5. 585–614.

Cheshire, Jenny, Paul Kerswill, Sue Fox and Eivind Torgensen. 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2): 151-196.

De Fina, Anna. 2011. Researcher and informant roles in narrative interactions: Constructions of belonging and foreign-ness. *Language in Society* 40(1): 27-38.

Duff, Patricia A. 2015. Transnationalism, multilingualism, and identity. *Annual Review of Applied Linguistics* 35. 577–80.

Espiritu, Yen Le & Thom Tran 2002. Viêt Nam, Nu'óc Tôi (Vietnam, My Country): Vietnamese Americans and transnationalism. In Peggy Levitt & Mary C. Waters, (eds.), *The Changing Face of Home: The Transnational Lives of the Second Generation*, 367–398. New York: Russell Sage Foundation.

Fehringer, Carol & Karen Corrigan. 2015. The rise of the *going to* future in Tyneside English: Evidence for further grammaticalisation. *English World Wide* 36(2). 198–227.

Fox, Sue. 2015. *The New Cockney: New Ethnicities and Adolescent Speech in the Traditional East End of London.* London: Palgrave Macmillan.

Hoffman, Michol F. & James A. Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22. 37–67.

Hundt, Marianne. 2014. Zero articles in Indian Englishes: A comparison of primary and secondary diaspora situations. In Marianne Hundt & Devyani Sharma (eds.), *English in the Indian Diaspora*, 131-170. Amsterdam: Benjamins.

Holliday, Adrian. 2010. Analysing qualitative data. In Brian Paltridge & Aek Phakiti, (eds.), *Continuum companion to research methods in applied linguistics*, 98–

110. London and New York: Continuum.

Joseph, John E. 2004. *Language and identity: National, ethnic, religious*. New York: Palgrave Macmillan.

Mendoza-Denton, Norma. 2002. Language and identity. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 475–499. Malden: Blackwell.

Nordmann, Rebekka. 2016. Rhoticity in the Indian English London Diaspora – A Cross-Generational Comparison of the Realization of Post-Vocalic /r/. Unpublished Research Paper, University of Zurich.

Phinney, Jean S. 1992. The Multigroup Ethnic Identity Measure: A New Scale for Use with Diverse Groups. In *Journal of Adolescent Research* 7(2). 156–176.

Phinney, Jean S. & Ong, Anthony D. 2007. Conceptualization and Measurement of Ethnic Identity: Current Status and Future Directions. *Journal of Counseling Psychology* 54(3). 271–281.

Rogers, Chandrika K. 2002. Syntactic features of Indian English: An examination of written Indian English. In Randi Reppen, Susan M. Fitzmaurice & Douglas Biber (eds), *Using Corpora to Explore Linguistic Variation*, 187–202. Amsterdam: Benjamins.

Schleef, Erik. Forthcoming. Identity and indexicality in the study of World Englishes. to appear in Daniel Schreier, Marianne Hundt and Edgar Schneider, eds. *Cambridge Handbook of World Englishes*. Cambridge: Cambridge University Press.

Sedlatschek, Andreas. 2009. *Contemporary Indian English: Variation and change.* Amsterdam and Philadelphia. John Benjamins.

Sharma, Devyani. 2005. Language transfer and discourse universals in Indian English

article use. *Studies in Second Language Acquisition* 27(4). 535–566.

Sharma, Devyani. 2011. Style repertoire and social change in British Asian English. *Journal of Sociolinguistics* 15(4): 464–492.

Sharma, Devyani & Lavanya Sankaram. 2011. Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change* 23. 399–-428.

Sharma, Devyani. 2012. English in India. In Alexander Bergs & Laurel Brinton (eds.), *Historical Linguistics of English*, Vol II, 2077–2091. Berlin: De Gruyter.

Spolsky, Bernard. 2016. *The Languages of Diaspora and Return*. (Special Issue of *Multilingualism and Second Language Acquisition* 1.2-3. Leiden: Brill.

Urban Dictionary, online, crowd-sourced dictionary of slang words and phrases, founded in 1999 by Aaron Peckham. (http://www.urbandictionary.com, last accessed 23.02.2015).

Williams, Ashley M. 2005. Constructing and reconstructing Chinese American bilingual identity. In James Cohen, Kara T. McAlister, Kellie Rolstad & Jeff MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism,* 2349–2356. Somerville, MA: Cascadiall Press.

Wodak, Ruth, Rudolf de Cillia, Martin Reisigl & Karin Liebhart. 1999. *The Discursive Construction of National Identity*. Edinburgh: Edinburgh University Press.

Zipp, Lena & Dellwo, Volker. 2011. 'Read speech normalization' (RSN): A method to study prosodic variability in spontaneous speech. In Wai-Sum Lee and Eric Zee, eds. *Proceedings of the 17th International Congress of Phonetic Sciences*, 17-21 August 2011, Hong Kong, 2328-2331.

AUTHOR. 2016. English in San Francisco Chinatown. Indexing identity with speech

rhythm? In Elena Seoane & Cristina Suárez-Gómez (eds.), *World Englishes: New Theoretical and Methodological Considerations*, 205–228. Amsterdam: Benjamins.

Table 1. Composition of the sample ('H' and 'G' label the participants' first language, i.e. Hindi or Gujarati; the number indicates the order in which the recordings were made.)

	first language	gender	generation	counterpart	generation	relationship
H01	Hindi	female	2nd	female	2nd	sisters
H04	Hindi	male	1st	male	1st	son and father
G01	Gujarati	female	2nd	female	2nd	friends
G02	Gujarati	female	1st	female	2nd	friends
G03	Gujarati	female	2nd	female	2nd	sisters
G04	Gujarati	female	2nd	female	1st	daughter and mother

Table 2. Adapted Multigroup Ethnic Identity Measure (MEIM-R), source: Phinney and Ong (2007: 276)

Item no.	Item
1	I have spent time trying to find out more about my ethnic group, such as its history, traditions, and customs.
2	I have a strong sense of belonging to my own ethnic group.
3	I understand pretty well what my ethnic group membership means to me.
4	I have often done things that will help me understand my ethnic background better.
5	I have often talked to other people in order to learn more about my ethnic group.
6	I feel a strong attachment towards my own ethnic group.

Table 3. Identity Scores – interview vs. questionnaire data (ranging from 1 most integrated into the 'host' community to 5 most 'ethnic' for both sets of data)

Participant (generation)	First rating (A1/A2)	Agreed rating	Questionnaire data	Difference
H01 (2nd)	2.0/2.6	2.5[11]	1.8	0.7
G01 (2nd)	(jointly rated from the outset)	2.8	2.5	0.3
G03 (2nd)	2.6/3.75	3.2	2.7	0.5
G04 (2nd)	2.3/2.35	2.4	2.0	0.4
H04 (1st)	1.8/2.2	1.9	2.8	1.3
G02 (1st)	1.8/2.39	2.0	1.2	0.8

Table 4. Variable article use (first-generation, baseline participant)

	Map Task	Discussion
all articles	40	17
all null	5	14
% null	11.1	45.2

Table 5. Variable article use (article: null article; % null articles) among the second generation participants

	MT in-group		MT out-group		Interview		Discussion	
		% null		% null		% null		% null
H01	118:16	11.9	75:4	5.1	46:2	4.2	18:0	0
G01	63:6	8.7	42:5	10.6	43:4	8.5	28:1	3.4
G03	100:11	9.9	52:4	7.1	155:5	3.1	62:2	3.1
G04	85:5	5.5	62:1	1.6	47:0	0	30:4	11.8
Total	366:38	9.4	230:14	5.7	291:11	3.6	138:7	4.8

Table 6. Vernacular feature density (frequency per thousand words) – first generation baseline speaker

MT in-group	Discussion
43	47.7

Table 7. Vernacular feature density (frequency per thousand words) – second generation

	MT in-group	MT out-group	Interview	Discussion
H01	38.2	41.7	2.9	5.4
G01	34.7	21.7	4.1	1.3
G03	19.8	29.5	3.1	3.0
G04	18.5	8.7	1.4	3.2

Table 8. Vernacular features across different contexts (excluding null articles)

	MT in-group	MT out-group	Interview	Discussion
H01	28.4	35.5	2.2	3.3
G01	24.8	10.8	1.0	0.7
G03	14.2	22.9	1.6	1.0
G04	15.2	6.5	1.4	0.9

Table 9. Identity scores – interview, questionnaire and combined

	Identity score interview	Identity score questionnaire	Identity score (mean)
H01	2.5	1.8	2.2
G01	2.8	2.5	2.7
G03	3.2	2.7	3.0
G04	2.4	2	2.2

Table 10. Vernacular features across different contexts (including null articles)

	MT in-group	MT out-group	Interview	Discussion
H01	38.3	41.7	2.9	6.0
G01	34.7	21.7	4.1	2.0
G03	21	29.5	3.1	3.0
G04	19.2	8.7	1.4	3.2

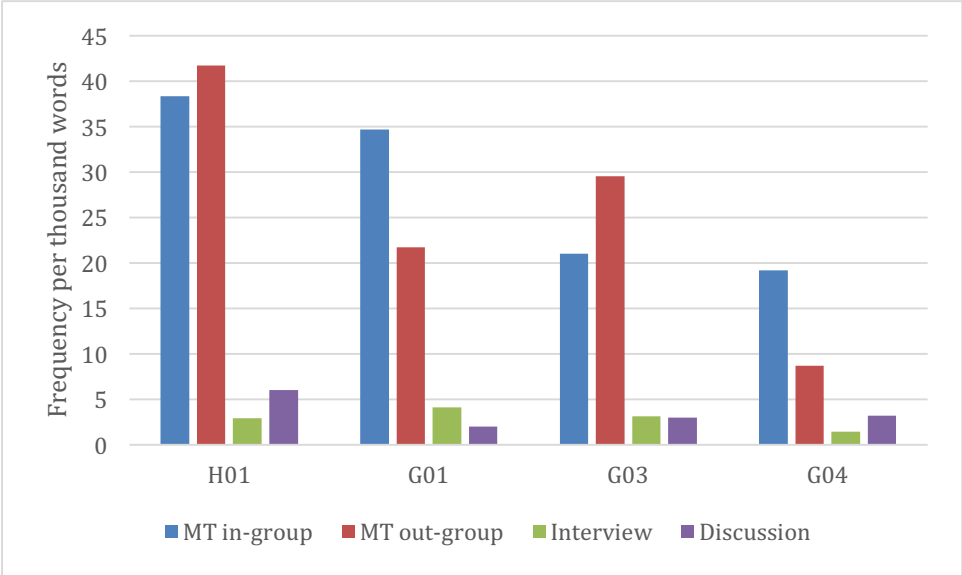

Figure 1. Vernacular features across different contexts (including null articles)